



# Praxisprojekte 2013

Projektberichte aus dem  
berufsbegleitenden Masterstudiengang  
*Bibliotheks- und Informationswissenschaft*  
der Fachhochschule Köln



Fachhochschule Köln  
Cologne University of Applied Sciences

Institut für Informationswissenschaft  
Institute of Information Science

2013





Band 44

**b.i.t.online – Innovativ**

Band 44

**MALIS-Praxisprojekte 2013**

Projektberichte aus dem berufsbegleitenden Masterstudiengang  
Bibliotheks- und Informationswissenschaft  
der Fachhochschule Köln

2013

Verlag: Dinges & Frick GmbH, Wiesbaden

## **MALIS-Praxisprojekte 2013**

Projektberichte aus dem berufsbegleitenden Masterstudiengang  
Bibliotheks- und Informationswissenschaft  
der Fachhochschule Köln

Herausgegeben  
von

ACHIM OßWALD  
INKA TAPPENBECK  
HAIKE MEINHARDT  
HERMANN RÖSCH

2013

Verlag: Dinges & Frick GmbH, Wiesbaden

## **b.i.t.online – Innovativ**

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;  
detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISBN 978-3-934997-51-6

ISBN 978-3-934997-51-6

ISSN 1615-1577

Redaktion: Dorothee Hofferberth und Susanne Röltgen

Satz: Dorothee Hofferberth

Titelfoto: © iStockfoto.com

© Dinges & Frick GmbH, 2013 Wiesbaden

Alle Rechte vorbehalten, insbesondere die des Nachdrucks und der Übersetzung.  
Ohne Genehmigung des Verlages ist es nicht gestattet, dieses Werk in einem  
photomechanischen oder sonstigen Reproduktionsverfahren zu vervielfältigen und  
zu verbreiten.

Alle Beiträge dieses Bandes werden auch als Open-Access-Publikationen über die  
Fachhochschule Köln sowie über den Verlag bereitgestellt.

Satz und Druck: Dinges & Frick GmbH, Wiesbaden

Printed in Germany

# **MALIS-Praxisprojekte 2013**

## **Projektberichte aus dem berufsbegleitenden Masterstudiengang Bibliotheks- und Informationswissenschaft der Fachhochschule Köln**

herausgegeben  
von

ACHIM OßWALD  
INKA TAPPENBECK  
HAIKE MEINHARDT  
HERMANN RÖSCH

Fachhochschule Köln  
Fakultät für Informations- und Kommunikationswissenschaften  
Institut für Informationswissenschaft

Vorwort	9
Einführung Inka Tappenbeck / Achim Oßwald	13
Vorbereitungen zur Langzeitarchivierung einer Fotokollektion Melanie Engels	17
Was kann eLib? Kommunikationsstrategie zur Einführung von Discovery-System, Linkresolver und Bibliothekssystem in den Fraunhofer-Bibliotheken Kerstin Hölscher / Cornelia Lang	35
Google Scholar versus EBSCO Discovery Service: Ein vergleichender Retrieval-Test Harald Kaluza	59
Immaterielle Leistungsanreize in Öffentlichen Bibliotheken: Situation und Maßnahmenkatalog Elisabeth Lappe-Oeynhausen	81
Virtuelle Lernräume an Hochschulen: Auswertung der Umfrage zur Neugestaltung von Lernräumen an Hochschulen der DINI AG Lernräume Anke Petschenka	107
Einführung eines RFID-basierten Selbstabholbereichs für Fernleihbestände an der ULB Düsseldorf Martin Reymer	121



Entwicklung von Schulungsmaterialien für Einsatzmöglichkeiten von MEI im bibliothekarischen Bereich Kristina Richts	137
Kleine Funktion, große Auswirkung: Die Funktionserweiterung „Zentrale Titelumlenkungen“ im hbz-Verbund Stephani Scholz	157
Forschungsdaten in den Osteuropawissenschaften: Ergebnisse einer Umfrage Doris Škarić	177
„Welcome to the library ...“: Englischsprachige Audio-Touren als technologiegestützte Informationsdienstleistung in wissenschaftlichen Bibliotheken Katrin Steiner	195
Informationskompetenz on demand: Individualisierte Rechercheleitfäden als Service für Schüler an der Hochschulbibliothek der RWTH Aachen Ina van Kann	215
Betreuerinnen und Betreuer der MALIS-Projekte: Kurzprofile	233

# **Vorbereitungen zur Langzeitarchivierung einer Fotokollektion**

Melanie Engels

## **Abstract**

Für die digitale Langzeitarchivierung von Bilddateien existieren derzeit bereits empfohlene Workflows und spezielle Tools. Im Rahmen dieses Projektes wurde der im nestor-Handbuch skizzierte Archivierungsworkflow für Bilddateien anhand der aus TIFF- und RAW-Dateien bestehenden Fotokollektion eines professionellen Fotografen unter Verwendung des Planungstools Plato erprobt. Der resultierende spezielle Workflow für TIFF-Dateien wird zukünftig für die Archivierung genutzt, während die RAW-Dateien vorerst nicht langzeitarchiviert werden.

For the long-term preservation of digital image files recommended workflows and special tools are already established. In this project the archiving workflow for image files outlined in the nestor-Handbuch was tested on a professional photographer's photo collection consisting of TIFF and RAW files, using the planning tool Plato. The resulting specialized workflow for TIFF images will be used for archiving those image files, while a long-time preservation of the RAW files has not been considered as appropriate.

## **1. Einleitung**

Die Langzeitarchivierung digitaler Bilddokumente wird im Kontext der fortschreitenden Digitalisierung analoger Bestände in Bibliotheken, Archiven und Museen zu einer weiteren Aufgabe für die genannten Einrichtungen. Vielfach ist die Digitalisierung von Büchern, Handschriften, Archivgut und anderem Kulturgut dadurch motiviert, dass die Nutzung des analogen physischen Exemplars dem konservatorischen Erhalt entgegensteht, die Nutzung also das Buch schädigt und man durch die Digitalisierung des Kulturguts den Erhalt des physischen Exemplars verbessern kann bei gleichzeitiger Verbesserung des Zugriffs auf das Werk an sich, da dieses nun nicht nur in Sonderlesesälen eingeschränkt genutzt werden kann, sondern in digitaler Form zur allgemeinen Nutzung bereitsteht.

Ebenso wie nun aber das physische Exemplar gepflegt und konserviert wird, muss auch das digitale Objekt gepflegt und sein Erhalt sichergestellt werden. Für physische Exemplare wie Bücher und Handschriften sind in der Regel bereits spezielle Erhaltungsstrategien definiert, die bestimmte Geschäftsgänge und Verfahren beinhalten, z.B. hinsichtlich Lagerung, Kontrolle und Zugriff. Ähnliche organisatorische Abläufe für die Langzeitarchivierung digitaler Objekte definiert das OAIS-Referenzmodell.<sup>1</sup> Die jeweiligen Ausprägungen hinsichtlich des zu archivierenden Objekttyps werden derzeit in vielen best-practice-orientierten Projekten evaluiert.

Aber nicht nur die BAM-Einrichtungen beschäftigen sich mit der langfristigen Archivierung digitaler Objekte. Auch in der Privatwirtschaft müssen digital vorliegende Informationen und Objekte archiviert werden, besonders wenn sie das Kapital eines Unternehmens ausmachen. Ein Beispiel hierfür ist die Berufsgruppe der Fotografen, deren Arbeitsumfeld sich in den letzten zehn Jahren durch die Wende von der analogen zur beinahe ausschließlichen digitalen Fotografie erheblich gewandelt hat. Hatten Fotografen zur Zeit der analogen Fotografie noch ein großes Spezialwissen hinsichtlich der verwendeten Materialien wie Fotopapier und Entwickler und der bestmöglich Archivierungsumgebung für ihre Negative, so müssen sie sich heute – wie viele andere Berufsgruppen auch – mit Speichermedien, Dateiformaten und technischen Workflows beschäftigen. Die Bilder entstehen in der Regel nur noch digital und werden auch digital weiterverarbeitet. Daher müssen sie auch digital gespeichert bzw. langfristig archiviert werden. Ein Verlust der Bilder kommt bei einem Fotografen einem Verlust seines Kapitals gleich. Daher wurde in diesem Projekt am Beispiel der Fotokollektion eines Fotografen ermittelt, welche Tools zur Umsetzung einer OAIS-konformen Archivierung von digitalen Bildern genutzt werden können und wie hoch der Aufwand in Bezug auf die Langzeitarchivierung der gesamten Kollektion eingeschätzt werden muss.

## 2. Zielsetzung und Konzeption

Ziel dieses Projekts<sup>2</sup> war es, im Sinne einer Machbarkeitsstudie zu überprüfen, welche Tools und Strategien für die Langzeitarchivierung dieser speziellen Fotokollektion genutzt werden können, um auf dieser Grundlage eine Aufwandsabschätzung für die Archivierung der gesamten bestehenden Kollektion vornehmen zu können. Dabei mussten das spezielle Profil der vorhandenen Kollektion, die vorhandenen Speichermedien, die technisch-organisatorischen Anforderungen und die Anforderungen des Kollektionsinhabers beachtet werden. Hierbei sollte auf das OAIS-Modell Bezug genommen werden, um eine transparente Planung und vertrauenswürdige Archivierung zu gewährleisten.

---

1 Open Archival Information System

2 Das Projekt erfolgte im Rahmen des berufsbegleitenden Masterstudiengangs Bibliotheks- und Informationswissenschaft an der FH Köln. Betreuer war Prof. Dr. Achim Oßwald.

## 2.1 Profil der Kollektion

Bei der zu archivierenden Kollektion handelt es sich um gescannte Bilder und Born-digital-Bilder einer Nikon D80.

Die gescannten Bilder wurden mit Hilfe eines Plustek Negativ-Scanners digitalisiert und teilweise mit Photoshop CS5 nachbearbeitet. Sie liegen im TIF-Format vor und sind analog der Archivierung der Negative benannt und abgelegt. Die Gesamtzahl der gescannten Bilder beläuft sich auf ca. 600 Dateien. Die durchschnittliche Dateigröße liegt zwischen 8 MB und 70 MB.

Die gescannten Bilder enthalten neben den Basismetadaten wie Breite und Höhe nur wenige eingebettete technische Metadaten im Exif-Standardformat und reduzierte administrative Metadaten. Deskriptive Metadaten wurden bisher nicht vergeben. Die Dateinamen orientieren sich an der Namensgebung für die analogen Negative, vielfach ist dies nur die Angabe von Aufnahmeort und Aufnahmejahr, z.B. „Irland 2004“.

Die Born-digital-Bilder liegen im NEF-Format vor, dem proprietären RAW-Format des Kameraherstellers Nikon. Beim RAW-Format handelt es sich – einfach ausgedrückt – um das digitale Negativ, das mit Hilfe eines RAW-Konverters wie im Fotolabor zum endgültigen Bild entwickelt werden kann. In der RAW-Datei liegen nur die vom Sensor der Kamera erkannten Grauwerte vor, die dann durch den Algorithmus des RAW-Konverters entsprechend interpretiert werden. Der Fachausdruck hierfür ist Interpolation.<sup>3</sup>

Einige Dateien weisen hierbei bereits Filialdokumente im von Adobe eingeführten offenen Metadatenstandard XMP auf, die durch Nachbearbeitungsprozesse in Photoshop CS5 entstanden sind. Aufgrund der Tatsache, dass nur zu einigen Bilddateien XMP-Dateien vorliegen, kann die Anzahl der Bilder nicht exakt bestimmt werden, es handelt sich schätzungsweise um ca. 25.000 Bilddateien.

In den RAW-Dateien sind bereits umfangreiche technische Metadaten im Exif-Format enthalten, wie z.B. die Blende, ob der Blitz ausgelöst wurde oder nicht, der ISO-Wert etc., also im Grunde alle Werte, die von der Kamera gemessen werden können. Diese können je nach Ausstattung der Kamera verschieden umfangreich sein. Zu den administrativen Metadaten können neben dem Aufnahmezeitpunkt auch der Zeitpunkt der letzten Bearbeitung etc. gehören oder die Angabe des jeweiligen Bearbeitungsprogramms, sowie Kamerahersteller, Modell usw.

---

3 Vgl. Gulbins; Steinmüller 2011, S. 6 ff.

## 2.2 Speichermedien

Als Speichermedium für die Fotokollektion diente bisher eine externe Festplatte mit 500 GB Speicher. Allgemein werden externe Festplatten nicht als langzeitarchivierungstauglich angesehen, da die geschätzte Lebensdauer dieses Mediums bei fünf Jahren liegt, von Langzeitarchivierung über einen Zeitraum von zehn Jahren hinaus also keine Rede sein kann.<sup>4</sup> Andererseits kann diese beschränkte Haltbarkeit des Speichermediums aber in den laufenden Archivierungsprozess mit einbezogen werden: Eine vernünftige Langzeitarchivierung plant auch immer in bestimmten zeitlichen Abständen eine Überprüfung ein, bei der abgewogen wird, ob eventuell eine Migration notwendig ist. Somit könnte die beschränkte Halbwertszeit des Mediums von ca. fünf Jahren dazu genutzt werden, alle fünf Jahre sowohl das Speichermedium als auch die Daten an sich hinsichtlich aktueller Formate und eventuellem Migrationsbedarf zu überprüfen.

Eine mithilfe des auch in diesem Projekt genutzten Planungstools Plato durchgeführte Studie, in der die Ansprüche verschiedener Typen von Fotografen an die Speichermedien evaluiert wurden, kommt zu dem Ergebnis, dass sich für professionelle Fotografen die externe Harddisk am besten eignet.<sup>5</sup>

## 2.3 Organisatorisch-technische Anforderungen

Um die Planung der Langzeitarchivierung möglichst transparent und nachvollziehbar durchzuführen, soll für den Planungsprozess das speziell für die Langzeitarchivierung entwickelte Tool Plato eingesetzt werden, das den Planungsworkflow steuert, sämtliche Planungsschritte und Entscheidungen dokumentiert und über implementierte Services zur Formaterkennung und Metadatenextraktion sowie über eine Auswahl von Migrationswerkzeugen verfügt. Dabei lehnt es sich eng an die im OAIS-Modell definierten Vorgaben für die Langzeitarchivierungsplanung an.<sup>6</sup>

Für den Archivierungsworkflow wurden die im nestor-Handbuch beschriebenen Schritte für die Behandlung digitaler Bilddokumente gewählt, die vor der Archivierung der Datei auch die Anreicherung mit zusätzlichen Metadaten sowie die Auslagerung einer Metadatendatei als Filialdokument für den objektunabhängigen Zugriff auf die Metadateninformationen vorsehen.<sup>7</sup> Diese Aufgaben werden als projektstrukturierende Arbeitsschritte in Kapitel 3 näher erläutert.

---

4 Vgl. Däßler 2010, S. 17.

5 Vgl. Becker; Rauber 2007, S. 72 ff.

6 Vgl. Becker 2010.

7 Vgl. Enders 2010, S. 18.

## 2.4 Anforderungen des Kollektionsinhabers

Für die Planung des Verfahrens mussten auch die Vorgaben des Kollektionsbesitzers beachtet werden, die vorsehen, dass alle im TIF-Format vorliegenden Dateien erhalten bleiben sollen, da dieses Format bereits als Standard für die Nach- und Weiterverarbeitung der Bilder implementiert ist. Außerdem wurde gewünscht, dass der Workflow möglichst mit Hilfe der Bildbearbeitungssuite Photoshop CS5 gestaltet wird und keine zusätzlichen Bildverarbeitungsprogramme, gleich ob Freeware oder kostenpflichtig, genutzt werden. Da der Kollektionsinhaber sich selber als eher durchschnittlich technikaffin charakterisierte, sollte darauf geachtet werden, dass die zu verwendenden Tools über eine grafische Benutzeroberfläche verfügen und nicht kommandozeilenbasiert sind.

## 2.5 Bildverwaltung

Die Verwaltung der archivierten Bilder erfolgte bisher ausschließlich über das Dateisystem. Eine Suche nach Metadaten, die Anzeige von Vorschaubildern für RAW-Dateien etc. war nicht möglich. Innerhalb der Langzeitarchivierungsplanung sollte deshalb auch der im OAIS-Modell definierte Aspekt Access beachtet werden.<sup>8</sup> Da der Kollektionsinhaber die Verwendung von Tools aus dem Funktionsumfang von Photoshop CS5 ausdrücklich wünschte, wurde das zu den Ad-hoc-Bildbrowsern<sup>9</sup> zählende Tool Adobe Bridge innerhalb dieses Projekts sowohl für die Bildverwaltung als auch für die Anreicherung der eingebetteten Bilddateien mit deskriptiven Metadaten im IPTC-Format<sup>10</sup> genutzt werden.

# 3. Projektplanung

## 3.1 Strategie und geeignete Tools

In der Projektplanungsphase wurden zuerst mit Hilfe von Ratgeberliteratur und unter Bezug auf publizierte Studien und Praxisberichte geeignete Tools und Strategien für das Projektvorhaben recherchiert. Sehr hilfreich waren hierbei die im nestor-Handbuch

---

<sup>8</sup> Vgl. Brübach 2010, S. 11.

<sup>9</sup> Ad-hoc-Bildbrowser erlauben das Browsen in / auf Festplatten / Datenträgern innerhalb einer Ordnerstruktur abgelegten Dateien; vgl. Gulbins; Steinmüller 2011, S. 578 f.

<sup>10</sup> IPTC-NAA ist ein Standard des International Press Communications Council zur Speicherung von Informationen zu Bildinhalten.

beschriebenen Vorgehensweisen für den Objekttyp Bild.<sup>11</sup> Wie bereits weiter oben erläutert, wird zu jeder Datei eine METS-Datei erstellt, in der die eingebetteten Bildinformationen extern gespeichert werden. Dadurch wird sichergestellt, dass auch bei Verlust oder Beschädigung der Bilddatei Zugriff auf diese Informationen besteht und diese für eine Wiederherstellung der Datei genutzt werden können. Aus der im nestor-Handbuch gegebenen Beschreibung wurde der für die praktische Umsetzung angestrebte Workflow abgeleitet. Die Schritte werden im Folgenden genauer definiert und die zur Ausführung des jeweiligen Schrittes recherchierten Tools genannt.

#### *Einfügen der deskriptiven Metadaten in das Bilddokument*

Bei diesem Vorgang werden deskriptive Metadaten wie Schlagwörter, Bildbeschreibung, Urheber etc. zur Bilddatei hinzugefügt. Hierfür kann der aus der Pressefotografie stammende IPTC-Standard genutzt werden, wie es auch in einem Langzeitarchivierungsprojekt des Bibliotheksservicezentrums Baden-Württemberg praktiziert wird.<sup>12</sup> Diese Daten können zum Beispiel mithilfe des Bildverwaltungstools Adobe Bridge geschrieben werden.

*Validieren des Datenformates des Bilddokuments / Extrahieren der Formatinformation (JHOVE) inkl. der Formatbestimmung (DROID) / Extrahieren der allgemeinen technischen Metadaten (Checksummen) / Generierung der technischen und Herkunftsmetadaten (MIX und PREMIS) aus den Formatinformationen*

Diese vier Schritte lassen sich nach den Erkenntnissen der bisherigen Recherchen und Tests zusammenfassen: JHOVE und FITS sind als Formatinformationstools ähnlich wie DROID bereits in das Planungstool Plato integriert. Beide Tools enthalten die Formatbestimmung durch DROID; außerdem ist FITS in der Lage, die Formatdaten und technischen Metadaten direkt im MIX-Format auszugeben. Auf die Bestimmung der Herkunftsmetadaten in PREMIS wurde im Projektverlauf verzichtet, da bei den in Phase zwei bearbeiteten gescannten Bildern nur geringe Herkunftsmetadaten anfallen, die auch in MIX erfasst werden können.

#### *Einfügen der technischen und Herkunftsmetadaten in ein Containerformat des Repositoriums*

In diesem Schritt sollen sämtliche aus der Bilddatei gewonnene Metadaten in eine METS-Datei geschrieben werden, die als Filialdokument gespeichert wird. METS ist in diesem Fall zwar nicht das Containerformat des Repositoriums, hat sich aber als Standard etabliert und erlaubt so auch die spätere Umwandlung in ein anderes Format, falls für die Archivierung einmal ein spezielles Repository-System genutzt werden sollte. Mit dem von der University of Florida entwickelten SobekCM METS Editor lässt sich eine Basis-METS-Datei erstellen, in die dann mit Hilfe eines XML-Editors die MIX-Daten eingefügt werden können.

---

11 Vgl. Enders 2010, S. 18.

12 Vgl. Schweibenz; Wolf 2010, S. 15.

### 3.2 Planung der Projektphasen

Wie bereits in den technisch-organisatorischen Anforderungen beschrieben, sollte zusätzlich zur Workflowplanung eine konkrete Planung des Archivierungsvorhabens mit Hilfe des Tools Plato durchgeführt werden. Diese Planung wurde als Phase eins der praktischen Umsetzung vorangestellt, da es mit Hilfe von Plato auch möglich ist, sowohl Tools zur Formaterkennung an Beispieldateien zu erproben als auch eventuell erforderliche Konvertierungsprogramme zu ermitteln.

Für die Phase zwei war die praktische Umsetzung des bereits skizzierten Ingest-Workflows für Bilddateien, am Beispiel von ca. 100 Beispieldateien mithilfe der in der ersten Phase durch Plato ermittelten Tools geplant. In der ursprünglichen Planung sollten als Testobjekte je 100 TIFF-Dateien und RAW-Dateien dienen. Bedingt durch Schwierigkeiten bei der Durchführung der Langzeitarchivierungsplanung mit Plato wurden nur 25 TIFF-Dateien bearbeitet, auf die Bearbeitung der RAW-Dateien wurde aufgrund der Evaluierungsergebnisse verzichtet.

Die einzelnen Aufgaben innerhalb der Phase zwei entsprechen dem oben skizzierten Ingest-Workflow.

## 4. Umsetzung

Im Folgenden wird nun die konkrete Umsetzung der im vorigen Kapitel vorgestellten Planung beschrieben. Auch die Schwierigkeiten bei der Umsetzung, die zu der bereits erwähnten Änderung der Planung hinsichtlich Bearbeitungszeit für die einzelnen Phasen und Anzahl der Testobjekte führten, werden dokumentiert.

### 4.1 Phase eins: Planung mit Plato

#### 4.1.1 Kurzprofil Plato

Für die Planung des gesamten Langzeitarchivierungsvorhabens wurde das Tool Plato ausgewählt, das von Wissenschaftlern der TU Wien im EU-Projekt PLANETS entwickelt wurde. Plato ist eine frei zugängliche, auf Java basierende Webapplikation, mit der der gesamte Planungsworkflow für ein Langzeitarchivierungsvorhaben schrittweise durchlaufen werden kann. Das Tool kann über die Plato-Website<sup>13</sup> genutzt werden. Derzeit steht Plato in der Version 3.0.1 bereit (Oktober 2010).

---

13 <http://www.ifs.tuwien.ac.at/dp/plato/intro.html>



Um Plato nutzen zu können, muss der Nutzer zuerst einen Account anlegen, um Demopläne ansehen zu können und eigene Pläne zu erstellen.

Bei der Erstellung eines eigenen Plans wird man automatisch schrittweise durch die einzelnen Planungsphasen geführt, ein Überspringen eines Schritts ist nicht möglich. Die einzelnen Phasen und Schritte innerhalb des Plato-Workflows werden nachfolgend an Hand der Langzeitarchivierungsplanung für TIFF- und RAW-Dateien dargestellt.

#### 4.1.2 Plato-Planungsworkflow für TIFF-Dateien

Ziel dieses Planungsworkflows sollte es sein, geeignete Tools für die in Phase zwei folgende Erstellung der externen Metadatendatei zu ermitteln. Dies schien auf Grund der Tatsache, dass das Formaterkennungstool DROID und die Metadatenerkennungstools JHOVE und FITS bereits als Webservice in Plato integriert sind, mit Hilfe von Plato möglich. JHOVE und FITS sind in der Lage, eingebettete Exif-Daten und IPTC-Daten zu erkennen und auszugeben.

##### *Define requirements / define basis*

Die ersten Planungsschritte in Plato finden im Abschnitt *define requirements* statt. Hier wurden zuerst die zur Identifikation nötigen Daten eingegeben wie Name des Plans, ID etc.

##### *Define requirements / define sample records*

Im nächsten Schritt wurde die Kollektion selber näher definiert, z.B. MIME-Type, Anzahl der Objekte und Zuwachsrate und einige Beispieldateien hochgeladen, die durch die in Plato als Webservice integrierten Tools direkt validiert wurden. DROID übernimmt hierbei die Formaterkennung. Zusätzlich zu dieser automatischen technischen Beschreibung der Beispieldateien wurde noch eine kurze intellektuelle Beschreibung hinzugefügt, in der auch der Auswahlgrund festgehalten wurde, nämlich sowohl große Dateien (ca. 45 MB) als auch kleinere Dateien (ca. 25 MB) zu evaluieren, um abschätzen zu können, ob sich dies auch auf die Performanz der Tools auswirkt.

##### *Identify requirements / objective tree*

In der nächsten Workflowphase war nun der sogenannte *objective tree* zu bestimmen, der die Anforderungen aller Beteiligten in Form einer Baumstruktur erfassen soll.<sup>14</sup>

Hierzu wurde als Basis ein bereits bestehender Strukturbaum für gescannte Bilder ausgewählt, der dann entsprechend der erarbeiteten Anforderungen von Kollektionsinhaber und Projektbeauftragter modifiziert wurde. Die grundlegende Frage hierbei lautete, welche Eigenschaften der Objekte von den auszuwählenden Tools in jedem Fall erkannt werden müssen, sowohl was die Beschaffenheit des Objekts an sich als auch die vorhandenen Metadaten angeht. Außerdem wurden prozessorientierte Anforderungen wie

---

14 Vgl. Becker 2010, S. 8 f.

Geschwindigkeit und Nutzerfreundlichkeit definiert sowie der Aufwand für das Setup.

Die *significant properties*, also die objektdefinierenden Eigenschaften, wurden aufgeteilt in die Bereiche *content* und *context*. *Content* bezieht sich dabei auf die tatsächlich das Objekt definierenden Eigenschaften, die bei einer Migration unbedingt erhalten bleiben müssen, während *context* sich auf die eingebetteten Metadaten bezieht. Zur Definition der objektdefinierenden Eigenschaften wurde auf die Ergebnisse einer Studie innerhalb des InSPECT-Projekts zurückgegriffen, deren Ziel es war, diejenigen Eigenschaften von Rastergrafiken zu bestimmen, die für jede Rastergrafik gegeben sind und das Wesen des Objekttyps ausmachen. In jener Studie wurden die folgenden *significant properties* ermittelt:<sup>15</sup>

- *Image width / image height*: Breite und Höhe des Bildes
- *XResolution / YResolution*: Auflösung des Bildes auf der X- und Y-Achse
- *Bits per sample*: Bits pro Farbkomponente
- *Samples per pixel*: Anzahl der Farbkomponenten pro Pixel

Im Bereich *context* wurden bestimmte Exif- und IPTC-Metadaten definiert, die von den Tools unbedingt erkannt werden müssen. Im Bereich *Nutzerfreundlichkeit* wurde die Dokumentation des Tools sowie die direkte Ausgabe der Daten im MIX-Format als Anforderungen definiert. Bei allen bisher angegebenen Anforderungen handelt es sich um Bool'sche Werte, die mit Ja / Nein besetzt sein können.

Im Abschnitt *Prozess* wurden dann auch andere Werte mit einbezogen, z.B. die Geschwindigkeit in MB/sec. Formatvalidierung und Batchprozessfähigkeit wurden wiederum als Bool'sche Werte definiert. Der Aufwand für Konfiguration / Vorbereitung wurde in Minuten pro Bild gewertet.

#### *Evaluate alternatives*

In dieser Phase des Planungsworkflows wurden die Tools ausgewählt, die für die jeweilige Archivierungsaktion miteinander verglichen werden sollen. Plato bietet hier die Möglichkeit auf eine Reihe vorhandener Tools zurückzugreifen, die als Webservice in Plato integriert sind und im weiteren Planungsverlauf auch automatisch getestet werden können, z.B. diverse ImageMagick-Migrationsroutinen (Wandlung von TIFF in JPEG2000 etc.). Allerdings liegt hierbei der Fokus auf Migrationstools – DROID, JHOVE und FITS können hier nicht ausgewählt werden. Daher wurden diese Alternativen manuell eingetragen – mit dem Nachteil, dass die entsprechenden Tests außerhalb von Plato durchgeführt werden müssen und die Ergebnisse nicht automatisch durch Plato gemessen werden können. Für das Projekt wurde auf die in Plato integrierten Webservices zurückgegriffen mit der Vorgabe, bei positiver Entscheidung für die Umsetzung des Langzeitarchivierungsvorhabens lokale Installationen der Services mit professioneller Hilfe entsprechend umzusetzen.

---

15 Vgl. Montague 2010, S. 13.

### *Evaluate experiments*

Für die Evaluation wurde auf die im Abschnitt *define requirements / define sample records* bereits vorhandenen Erkennungsergebnisse der Tools DROID, JHOVE und FITS zurückgegriffen. DROID konnte als reines Formaterkennungstool allerdings kaum ein Kriterium erfüllen, hätte also gar nicht evaluiert werden müssen.

Die unter *content* aufgeführten *significant properties* wurden von JHOVE und FITS gleichermaßen erkannt. Auch die unter *context* aufgeführten Eigenschaften wurden erkannt und ausgegeben. Die unter *Prozess* und *Aufwand* definierten Eigenschaften konnten zum größten Teil nicht evaluiert werden und wurden deshalb für alle Tools gleichsam mit 0 bewertet.

### *Analyse results*

Die Ausgabe der Analyseergebnisse zeigte eine eindeutige Empfehlung für FITS an, das mit 5.00 die höchste Wertung erzielte; dahinter folgte JHOVE, weit abgeschlagen aufgrund der reinen Funktionalität als Formaterkennungstool DROID. In der *conclusion* wurde dokumentiert, dass FITS JHOVE darin übertrifft, dass es in der Lage ist, automatisch die gewünschten Werte im MIX-Format auszugeben. Erhoffter Effekt ist eine schnellere Erstellung der externen Metadaten-Datei, da die technischen Metadaten per Copy & Paste eingefügt werden können. Daher wurde FITS als Extraktionstool für die Phase zwei gewählt.

## **4.1.3 Plato-Planungsworkflow für RAW-Dateien**

Der Planungsworkflow für die RAW-Dateien unterscheidet sich vor allem dadurch von der TIFF-Planung, dass hierbei tatsächlich eine Migrationsstrategie untersucht wird, nämlich die Migration der vorhandenen RAW-Dateien in TIFF zur Langzeitarchivierung. In der Fachwelt ist es derzeit umstritten, ob sich RAW-Dateien für die Langzeitarchivierung eignen. Wie bereits erläutert, sind RAW-Dateien digitale Negative, d.h. man bewahrt nicht das fertige Objekt, sondern die Basis, aus der das Endprodukt erzeugt werden kann. Bei der Umwandlung in TIFF würde also diese Möglichkeit verloren gehen. Ein weiterer vorhersehbarer Nachteil ist die Dateigröße, die bei der Umwandlung von RAW in TIFF um ein Vielfaches vergrößert würde, da ja die durch den RAW-Konverter ergänzten Informationen zusätzlich gespeichert werden müssen.

### *Define requirements*

Die Anzahl der Bilder in der Kollektion (ca. 25.000 Bilder) und die jährliche Zuwachsrate (ca. 5.000 Bilder) zeichnen ein ganz anderes Profil der Kollektion, als es bei den zuvor evaluierten TIFF-Dateien der Fall war.

Eine Erkennung der Beispieldateien durch DROID, JHOVE und FITS scheitert, das For-

mat wird nicht erkannt. Die enthaltenen Exif- und IPTC-Daten werden aber korrekt ausgegeben. Allerdings ist es nicht möglich, die hochgeladenen Dateien in Plato zu verwenden; beim Speichern und Wechseln zum nächsten Schritt tritt ein Fehler auf.

### *Define requirements / identify requirements*

Der *objective tree* ähnelt dem für die TIFF-Dateien entworfenen Strukturbaum. Im Bereich *content* wurde allerdings die Sektion *picture taking conditions* eingefügt, da durch den Umstand, dass in diesem Fall nicht ein analoges Bild digitalisiert wird, sondern digitale Bilder direkt erzeugt werden, sehr viel mehr technische Metadaten zur Verfügung stehen. Auch bei den formatbezogenen Anforderungen wurden Eigenschaften ergänzt, z.B. die Verbreitung des Formats, die Kompression und eventueller Qualitätsverlust etc. Die prozessbezogenen Anforderungen und die relevanten Kostenfaktoren wurden beibehalten.

### *Evaluate alternatives*

In der Evaluierung zeigte sich direkt bei den objektbezogenen Eigenschaften, dass nach der Wandlung in das TIFF-Format die Werte für *height*, *width*, *XSampling* und *YSampling* nicht mehr übereinstimmten. Dies scheint aber in der Erkennung der NEF-Dateien durch JHOVE und FITS bedingt zu sein, da diese Werte in Adobe Bridge für TIFF- und NEF-Datei übereinstimmten. Trotzdem erweckte diese Abweichung Zweifel: Welche Werte stimmen, wieso treten bei der Erkennung durch FITS und JHOVE Abweichungen auf? Daher wurden hier die Werte aus Adobe Bridge statt der von FITS und JHOVE ermittelten Werte übernommen, was zu einer positiven Bewertung des Migrationsprozesses in diesem Bereich führte. Bei den Bits pro Farbkomponente zeigte sich eine Verringerung des Wertes von 16 Bit bei RAW zu 8 Bit bei TIFF. Auch bei den *samples per pixel* wich der Wert der TIFF-Datei von dem der RAW-Datei ab, weshalb dies ebenfalls negativ bewertet wurde.

Die IPTC-Daten blieben bei der Wandlung in TIFF erhalten. Auch die XMP-Daten konnten durch JHOVE und FITS in der TIFF-Datei erkannt werden.

Bei den formatbezogenen Anforderungen wurden NEF und TIFF in Bezug auf ihre Verbreitung, Dokumentation und Standardisierung bewertet; wie zu erwarten schnitt das proprietäre RAW-Datenformat NEF hier schlechter ab als der Quasi-Standard TIFF. Vorteile zeigten sich aber bei der Dateigröße: Im Durchschnitt sind die TIFF-Dateien fast dreimal so groß wie die ursprünglichen RAW-Dateien.

### *Analyse results*

Die Ergebnisse zeigten klar eine Empfehlung zugunsten der Beibehaltung der NEF-Dateien an. Ausschlaggebend hierfür sind die fehlenden bzw. fehlerhaften Exif-Daten in den TIFF-Dateien und der extreme Anstieg der Dateigröße bei Wandlung in TIFF.

Aufgrund dieser Ergebnisse wurde entschieden, die RAW-Dateien vorerst zu behalten und keine spezielle Langzeitarchivierungsmaßnahme zu ergreifen; stattdessen soll im

Sinne der *bitstream preservation* eine Pflege der vorhandenen NEF-Dateien vorgenommen werden, bis sich auf dem Feld der Archivierung von RAW-Dateien neue Erkenntnisse gewinnen lassen.

#### 4.2 Phase zwei: Anwendung des Planungsworkflows auf TIFF-Beispieldateien

Aufgrund der Ergebnisse der Workflowplanung mit Plato wurden in Phase zwei nicht wie ursprünglich vorgesehen TIFF- und RAW-Beispieldateien bearbeitet, sondern nur die TIFF-Dateien berücksichtigt.

Der bereits skizzierte Workflow wurde in der praktischen Anwendung folgendermaßen umgesetzt:

- Ergänzung der deskriptiven Metadaten im IPTC-Format mit Hilfe von Adobe Bridge
- Formatvalidierung mit DROID und Extraktion der MIX-Daten mit FITS via Plato Webservices
- Erstellung einer Basis-METS-Datei mit einfachen deskriptiven Metadaten in Dublin Core mit Hilfe des Tools SobekCM METS Editor, inklusive Checksummenerstellung für die Bilddatei
- Einfügen der MIX-Daten in die METS-Datei mit Hilfe des XML-Editors NotePad++
- Ablage der Bilddatei und der METS-Datei

Ziel dieser Phase war, den skizzierten Workflow zur Erstellung von *archival information packages* zu erproben. Die einzelnen Arbeitsschritte werden im Folgenden erläutert und hinsichtlich ihres Aufwands bewertet.

##### 4.2.1 Ergänzung der deskriptiven Metadaten

Im ersten Schritt des Workflows sollten die bestehenden eingebetteten Metadaten um deskriptive Metadaten im IPTC-Format ergänzt werden. Dies konnte mit Hilfe des Tools Adobe Bridge umgesetzt werden, entsprechend der Vorgabe des Kollektionsinhabers, wenn möglich auf die bestehenden Tools seiner Adobe Photoshop CS5 Installation zurückzugreifen.

Für die Stichwörter wurde überlegt, auf ein kontrolliertes Vokabular zurückzugreifen. In die engere Auswahl kam hierbei das von der zu Getty Images gehörenden Microstock-Agentur iStockphoto genutzte Vokabular, da der Kollektionsinhaber seine Bilder über diese Agentur anbietet. Allerdings stellte sich nach intensiver Recherche heraus, dass dieses Vokabular kaum dokumentiert ist und daher gar nicht außerhalb von iStockphoto

genutzt werden kann. Stattdessen wurde vom Kollektionsinhaber ein Basisvokabular entwickelt, das die von ihm bisher am häufigsten verwendeten Begriffe enthält und flexibel erweitert werden kann. Der zeitliche Aufwand der Metadatenerfassung pro Bild lag während des Tests bei ca. fünf Minuten, je nachdem wie umfangreich die Metadaten gestaltet wurden.

#### 4.2.2 Formatvalidierung und Metadatenextraktion

Da zur Formaterkennung und Metadatenextraktion keine lokalen Installationen der Tools DROID, JHOVE und FITS genutzt werden konnten, musste der Umweg über die in Plato integrierten Abfragen per Webservice genommen werden. Die Uploadzeit war dabei stark von der Dateigröße abhängig; für den Upload von Dateien größer 45 MB musste mit einer Uploadzeit von ca. fünf Minuten gerechnet werden. Diese Zeit konnte aber in der Regel mit der Bearbeitung anderer Bilder überbrückt werden, so dass hier kein übermäßig höherer zeitlicher Aufwand gegenüber der möglichen Batchverarbeitung bei lokaler Installation zu verzeichnen war.

#### 4.2.3 Erstellung einer Basis-METS-Datei

Für die Erstellung einer einfachen METS-Datei, in die die extrahierten technischen Metadaten im MIX-Format eingefügt werden können, wurde das von der University of Florida entwickelte und genutzte Tool SobekCM METS Editor in der Version 1.0.4 genutzt.<sup>16</sup> Hiermit ist die Erstellung einer METS-Datei inklusive deskriptiver Metadaten in Dublin Core oder MODS sowie gegebenenfalls die Anlage von Strukturinformationen möglich. Außerdem wird die referenzierte Bilddatei auf Wunsch mit einer Prüfsumme versehen und diese mit in die METS-Datei geschrieben. Mit Hilfe des METS Editors konnten die benötigten Basisdateien innerhalb von weniger als fünf Minuten erstellt werden; diese stellten eine gute Basis für die Ergänzung der technischen Metadaten im MIX-Format dar. Die integrierte Prüfsummenerstellung konnte zudem Zeit für Auswahl und Einsatz eines separaten Tools einsparen. Einzig die nötige manuelle Änderung der Dateiendung von .mets zu .xml verursachte unerwarteten zusätzlichen Aufwand.

#### 4.2.4 Einfügen der MIX-Daten

Für das Einfügen der MIX-Daten musste die erstellte Basis-METS-Datei mithilfe eines XML-Editors bearbeitet werden. Hierzu wurde der Editor Notepad++ in der Version 6.1.3 genutzt.

---

16 Vgl. METS Editor Version 1.0.4. <http://ufdc.ufl.edu/software/mets>

Mithilfe des Editors wurde aus der FITS-Dateianzeige in Plato der die MIX-Daten enthaltende Abschnitt kopiert und in die METS-Datei unterhalb des <METS:techMD>-Knotens eingefügt. Danach wurde die erzeugte Datei validiert und nach erfolgreicher Prüfung zusammen mit der Bilddatei gespeichert.

#### 4.2.5 Ablage der AIPs

Die erstellten *archival information packages* (AIP) wurden dann in dem für die Tests eingerichteten Archivbereich auf der externen Festplatte gespeichert. Dabei wurde auf die ursprüngliche Verzeichnisstruktur verzichtet; Bilddateien und METS-Dateien liegen alle gemeinsam in einem Verzeichnis, damit die Bildverwaltungskomponente Adobe Bridge dieses Verzeichnis schnell durchsuchen kann.

Durchschnittlich konnten die drei letzten Schritte *Erstellung der Basis-METS-Datei*, *Einfügen der MIX-Daten* und *Ablage der AIPs pro Bild* innerhalb von fünf Minuten durchgeführt werden.

Insgesamt muss also pro Bild – vorausgesetzt die Formatvalidierung und Metadatenextraktion kann im Hintergrund ausgeführt werden – mit einer Bearbeitungszeit von etwa zehn Minuten gerechnet werden: Fünf Minuten für die Erfassung der deskriptiven IPTC-Daten und fünf Minuten für die Erstellung der METS / MIX-Datei, inklusive Ablage.

### 5. Evaluation

#### 5.1 Allgemeine Vorbereitung

Die allgemeine Vorbereitung des Projekts mit Hilfe von Ratgeberliteratur zum Thema der Langzeitarchivierung von Bilddateien erlaubte bereits im Vorfeld einige der später benötigten Tools in ihrer Verwendung in anderen Projekten kennenzulernen, zu testen und zu bewerten. Daher wurden eher benutzerunfreundliche Tools, die nur auf Kommandozeilenebene zu bedienen sind, von vornherein ausgeschlossen. Der im nestor-Handbuch skizzierte Workflow für die Erstellung von AIPs konnte in der zweiten Projektphase praktisch unverändert umgesetzt werden. Durch die ausführliche Vorbereitung konnte das Projekt detailliert geplant werden; nur zum Planungstool Plato hätten eingehendere Recherchen und Tests im Vorfeld stattfinden müssen, um die schließlich in der Phase eins aufgetretenen Probleme vorher zu erkennen.

## 5.2 Phase eins

Die Planung mit Plato führte zwar in beiden Anwendungsfällen – TIFF-Dateien und RAW-Dateien – im Endeffekt zu einem ausführbaren Plan, hätte aber durch vorausgehende Recherchen oder Tests transparenter hinsichtlich des tatsächlichen Funktionsumfangs von Plato sein können.

Bei den TIFF-Dateien führte der Umstand, dass Plato zwar die Erkennung der Testdateien durch DROID, JHOVE und FITS unterstützt, diese aber nicht als Evaluationstools anbietet, zu einem zeitaufwändigen Workaround der außerdem nicht automatisch durch Plato gemessen werden konnte.

## 5.3 Phase zwei

Die in der zweiten Projektphase für die Langzeitarchivierung der TIFF-Dateien gewonnenen Daten ergeben, dass pro Bild ein Archivierungsaufwand von etwa zehn Minuten einzuplanen ist. Hochgerechnet auf die gesamte Kollektion von etwa 600 TIFF-Dateien ergibt dies einen zeitlichen Gesamtaufwand von 100 Stunden, also umgerechnet etwa 12,5 Personentage. Zu bewerten ist aber auch der Workflow und die Benutzerfreundlichkeit der verwendeten Tools. Die Eingabe der IPTC-Daten mithilfe von Adobe Bridge ist unkompliziert und wird inzwischen vom Kollektionsinhaber bereits bei der ersten Erfassung/Bearbeitung neuer Bilder geleistet.

Die Bedienung der übrigen Tools, z.B. des XML-Editors Notepad++, sind leicht erlernbar, da es sich hierbei im Grunde um bei jeder Bildbearbeitung zu wiederholende Schritte handelt.

## 6. Fazit

Insgesamt konnte das Projekt mit kleinen Einschränkungen wie geplant durchgeführt werden und führte zu den erwarteten Ergebnissen. Der Aufwand für die Erstellung der AIPs für die TIFF-Dateien wird vom Kollektionsinhaber als vertretbar akzeptiert. Als großer Erfolg dieses Projekts kann gewertet werden, dass der Kollektionsinhaber die Wichtigkeit deskriptiver und technischer Metadaten erkannt hat und mittlerweile für alle neu erstellten Bilder deskriptive Metadaten vergibt, teilweise im Batch, teilweise durch Autopsie. Hierzu trug maßgeblich bei, dass diese Anreicherung von deskriptiven Metadaten mit der in der vom Kollektionsinhaber bereitgestellten Softwaresuite Photoshop CS5 enthaltenen Software Bridge benutzerfreundlich vorgenommen werden kann und Bridge außerdem als Bildverwaltung die metadatenbasierte Suche



erlaubt. Vor Projektstart hatte der Kollektionsinhaber Bridge selten aktiv verwendet; die Erkenntnis der sinnvollen enthaltenen Funktionen kann also als ein verdienstvoller Nebeneffekt dieses Projektes gewertet werden.

Um die digitalen Bilddateien der untersuchten Fotokollektion langzeitarchivieren zu können, werden die folgenden Schritte empfohlen:

- Lokale Installation der genutzten Tools (FITS etc.)
- Bearbeitung der TIFF-Dateien nach Vorgabe des erprobten Workflows
- Redundante Speicherung der Kollektion und regelmäßige Überprüfung
- Überprüfung des zur Verschlagwortung genutzten Vokabulars.

Melanie Engels

studierte Linguistik und Germanistik an der Heinrich-Heine-Universität Düsseldorf und schloss ihr Studium 2009 mit der Masterprüfung ab. Seit 2010 arbeitet sie als Fachberaterin bei der ImageWare Components GmbH in Bonn. Neben ihrer beruflichen Tätigkeit absolviert sie seit März 2011 den Masterstudiengang Bibliotheks- und Informationswissenschaft der Fachhochschule Köln.

Kontakt: melanie.engels@smail.fh-koeln.de

## Literatur- und Quellenverzeichnis

Letztes Abrufdatum der folgenden Internet-Dokumente ist der 17.03.2013.

Becker, Christoph et al. 2010. Plato. In: nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Neuroth, Heike et al. (Hrsg.). Version 2.3. S. 3-20. urn:nbn:de:0008-20100305251

Becker, Christoph; Rauber, Andreas 2007. Langfristige Archivierung digitaler Fotografien. Wien. <http://www.ifs.tuwien.ac.at/dp/fotostudie/fotostudie.pdf>

Brübach, Nils 2010. Das Referenzmodell OAIS. In: Neuroth, Heike et al. (Hrsg.). nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. S. 3-14. urn:nbn:de:0008-2010061762

Däßler, Rolf 2010. Datenträger und Speicherverfahren für die digitale Langzeitarchivierung. In: Neuroth, Heike et al. (Hrsg.). nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3, S. 6-22. urn:nbn:de:0008-20100305200

Enders, Markus 2010. Bilddokumente. In: Neuroth, Heike et al. (Hrsg.). nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3, S. 8-18. urn:nbn:de:0008-20100305327

- Gulbins, Jürgen; Steinmüller, Uwe 2011. Handbuch digitale Dunkelkammer. Heidelberg: DPunkt Verlag.
- Montague, Lynne 2010. Significant Properties Testing Report: Raster Images. <http://www.significantproperties.org.uk/rasterimages-testingreport.pdf>
- METS Editor Version 1.0.4. <http://ufdc.ufl.edu/software/mets>
- Planets Preservation Planning Tool: Plato 3.0 User Manual V1.0. 2010 [http://www.ifs.tuwien.ac.at/dp/plato/docs/Plato\\_3\\_UserManual.pdf](http://www.ifs.tuwien.ac.at/dp/plato/docs/Plato_3_UserManual.pdf)
- Schweibenz, Werner; Wolf, Stefan 2010. Langzeitarchivierung eines digitalen Bildarchivs - Projekt zum Aufbau eines Langzeitarchivs für hochaufgelöste digitale Bilddateien der Staatsgalerie Stuttgart am BSZ. In: Neuroth, Heike et al. (Hrsg.). nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3, S.13-21. urn:nbn:de:0008-20100305416